

News Aggregator: The World at Your Finger Tips

Liyaqat Fayaz¹, Iqbal Bashir¹, Mrs. Sahila²

¹BE CSE, SSM College of Engineering, Jammu and Kashmir, India

²Assistant Professor, Department of CSE, SSM College of Engineering, Jammu and Kashmir, India

ABSTRACT

This aims to develop an online aggregator of news feeds and posts from different sources. News Aggregator is simply an online software which collects new stories and events around the world from various sources all in one place. News aggregator plays a very important role in reducing time consumption, as all of the news that would be explored through more than one website will be placed only in a single location. Also, summarizing this aggregated content absolutely will save reader's time. A proposed technique used called the Text Rank algorithm that showed promising results for summarization. This paper presents the main goal of this project which is developing a news aggregator able to aggregate relevant articles of a certain input keyword or key-phrase. Summarizing the relevant articles after enhancing the text to give the reader understandable and efficient summary.

KEYWORDS: News aggregator; text summarization, Web scraping, Web crawling, python

1. INTRODUCTION

It is a web application which aggregates data (news articles) from multiple websites. Then presents the data in one location. **News aggregator**, online platform or software device that collects news stories and other information as that information is published and organizes the information in a specific manner. News aggregation is based on the concept of content syndication, where content created by one or more news-gathering organizations is distributed through a different organization.

The biggest advantage of using a news aggregator website is that you get all your favourite news in one place. You don't have to visit all your favourite publications separately to read their latest content. All you have to do is visit your go-to news aggregator website and let it do the heavy-lifting for you.

A majority of these aggregator websites do not publish their own content. They simply fetch content from various publications using their RSS feeds and present them to you in a visually pleasing manner. This is why these websites are also sometimes referred to as RSS feed readers as well.

How to cite this paper: Liyaqat Fayaz | Iqbal Bashir | Mrs. Sahila "News Aggregator: The World at Your Finger Tips" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-6 | Issue-1, December 2021, pp.802-806, URL: www.ijtsrd.com/papers/ijtsrd47919.pdf



Copyright © 2021 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



1.1. WEB SCRAPING

Web scraping is the automated process of extracting data from web pages by using a robot (called a web crawler). Web scraping is data scraping used for extracting data from websites. The web scraping software may directly access the World Wide Web using the Hypertext Transfer Protocol or a web browser. While web scraping can be done manually by a software user, the term typically refers to automated processes implemented using a bot or web crawler. It is a form of copying in which specific data is gathered and copied from the web, typically into a central local database or spreadsheet, for later retrieval or analysis.

Web scrapers can be developed using programming languages using third party libraries. These libraries provide the HTTP connection with features like SSL certificates and authentication after establishing the connection parsing is done using third party libraries.

2. LITERATURE REVIEW

Decrease in Conventional News publication methods as stated by Brown, Jones, Patterson and Casero-Ripolles [3–6] the usage of printed and broadcasted news has decreased among young people. Mindich

[7] points out that 80% of America’s population under 30 do not retrieve news form newspaper on a daily basis, where 70% of American people above 30 do not as well. Furthermore Mindich [7] supports his statement by claiming that the median age of American people gather information from broadcasted sources is 60, hence hinting at a low value of young people watching news shows. Young people have various reasons that lead to such a decline: lack of time, preferences in different media or content that does not meet the people’s interests [6]. In addition to that Casero-RipollÁ’s [6] claims the the lack of relevance can be traced back to the missing connection to experiences and interest. In 2015 the Pew Research Center [8] has conducted a research, observing a decline in cable news compared to rise in the year 2013. Moreover they have observed that newspaper circualtion has declined from 2013 to 2014 [8]. In contrast Barnhurst,Wartella and Raeymaeckers [9, 10] claim that young people are generally interested in news, however the method of conventional spreading does not apply to them. As described by Raeymaeckers [10] young people find news in newspaper and on TV too difficult to understand. In addition to that Barnhurst and Wartella [9] approve CaseroRipolles [6] statement of the news not being reflective of young people’s lives. Nonetheless instead of stating that young people are not interested in political news and such, Reaymaeckers concludes that the news’ language does not fit with the youth [10]. They [10] propose that conventional news publisher and producer should

focus on helping the youth to understand the background and context of the news better. This conclusion gets supported by Meijer and Irene [11]. Meijer and Irene [11] state that news producers are required to develop a new of standard of news in order to appeal to young people. Finally it is possible to say that a decline in newspaper and broadcasted news is observable. Various factors play here an important role, nevertheless the method of translating news for the youth is the biggest issue. Young people do not feel connected to the news any more due to their presentation and language.

3. PROPOSED SYSTEM

Our proposed system works in two phases. Getting the preference of the user. Serving the personalised news on the news feed. User first logs in into the system, and chosés the news portals and category from the given options, then personalised news is served.

News aggregators scrape the data from different news portals in different categories and store them into a database. The objective of web scraping is to scrape the data from identified websites and convert it into a form which can be stored into traditional databases.

News aggregator works in three phases, first it scrapes the web for the news articles. Then it stores the image, link, and title of the article in the database. Then later the stored objects in the database are served to the news feed. The client gets information on his news feed.



FIG 1: USER CASE DIAGRAM

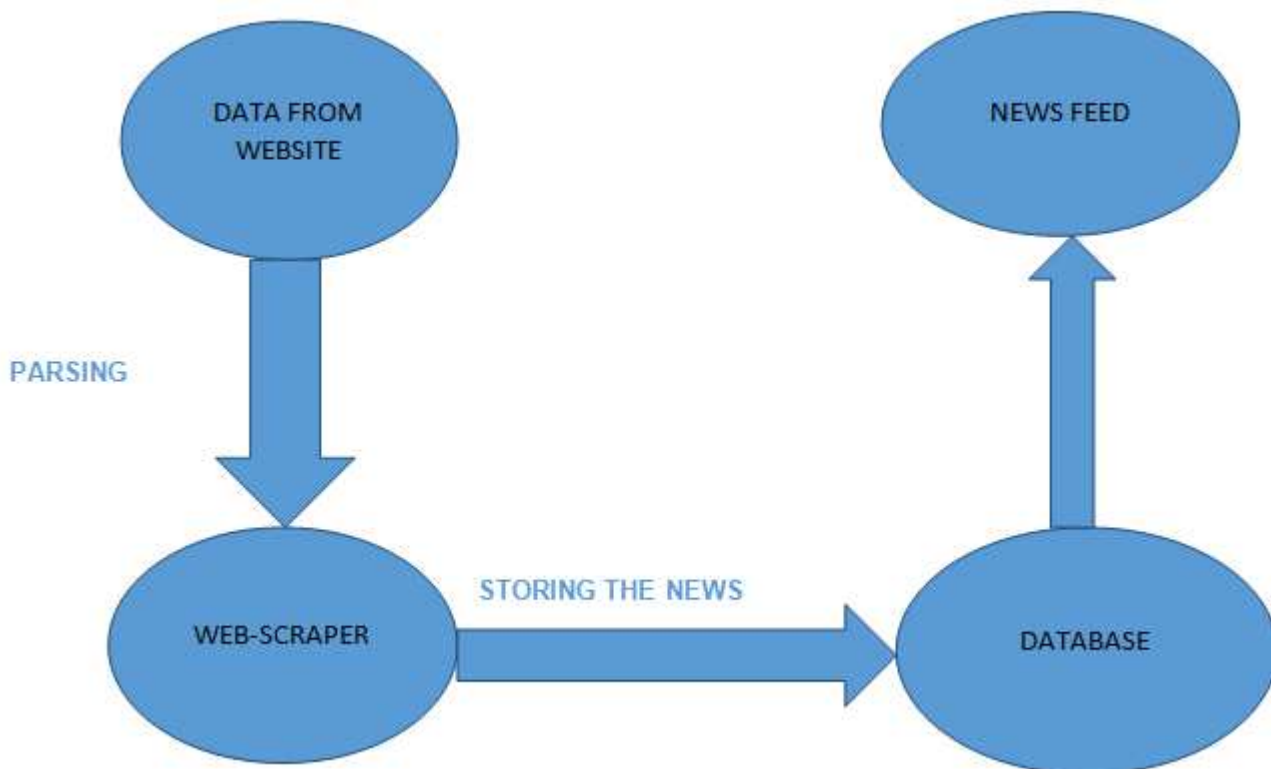


FIG 2: WORKING OF NEWS AGGREGATOR

4. METHODOLOGY:

- The following methodology was used in the news aggregator.
- Firstly, the URL of the webpage is required to be accessed to fetch the content. URLs were stored in a python dictionary
- For getting the HTML structure of the website, the HTTP connection has to be established to the web server.
- Web pages of news articles have a basic html structure, and which can be accessed and then required data can be fetched using python libraries.
- Then the scraped article was stored in the database.

The following system was created using python and Django

- To get started, python with the following libraries needs to be installed - requests, Django. And scrapy is used i.e. (Scrapy is an open-source and collaborative framework for extracting the data you need from websites in a fast and simple manner.)
- Next step is to create a web app to showing the news. Web apps have been developed using Django in the backend and HTML, CSS and JS in the frontend.
- Then the next step is to create the web scraper for the news portals. Since different websites have different structure, we cannot use the same scraper for all news portals, due to this different web scraper was created for the different websites.
- Then web scraper and Web app were integrated to make the system.
- Then the articles are filtered according to the user’s preference and rendered to the news feed.

We used the programming language i.e. python for the project. Python is a widely used general-purpose, high level programming language. It was created by Guido van Rossum in 1991 and further developed by the Python Software Foundation. It was designed with an emphasis on code readability, and its syntax allows programmers to express their concepts in fewer lines of code.

Python is a programming language that lets you work quickly and integrate systems more efficiently. It is an open source general purpose programming language with thousands of open source libraries for different range of works, which makes it easy to do a variety of work from web development to web scraping. Using these libraries made it easy to implement the system and reduced the size of code which made the code readable to other individuals. And because of these features python was chosen to implement this project.

The web app was built on Django web framework, which is an open source python web framework for rapid development of projects. A web framework is a tool consisting of components required for web app development. Django's inbuilt modules help in making the development process hassle free so one can focus on writing code without worrying about other side stuff.

5. SCREENSHOTS

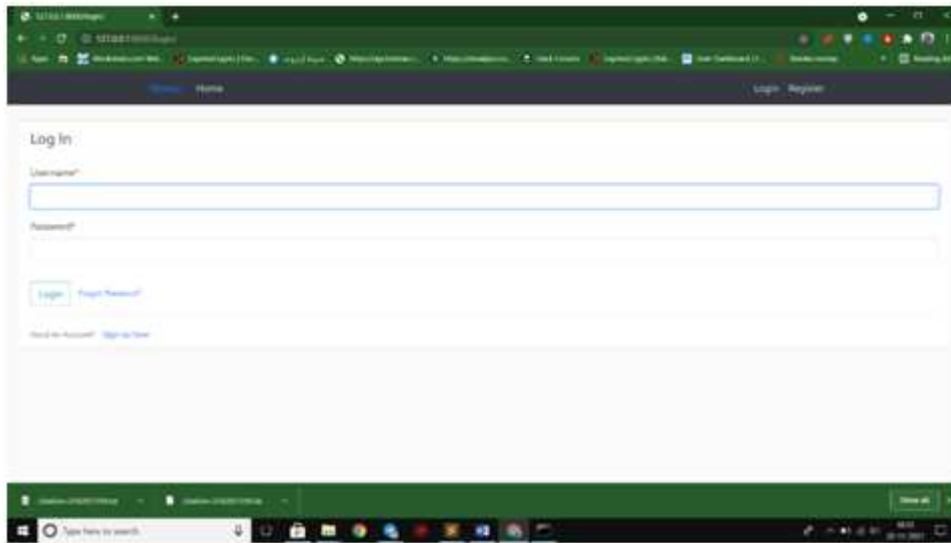


Fig 3: login

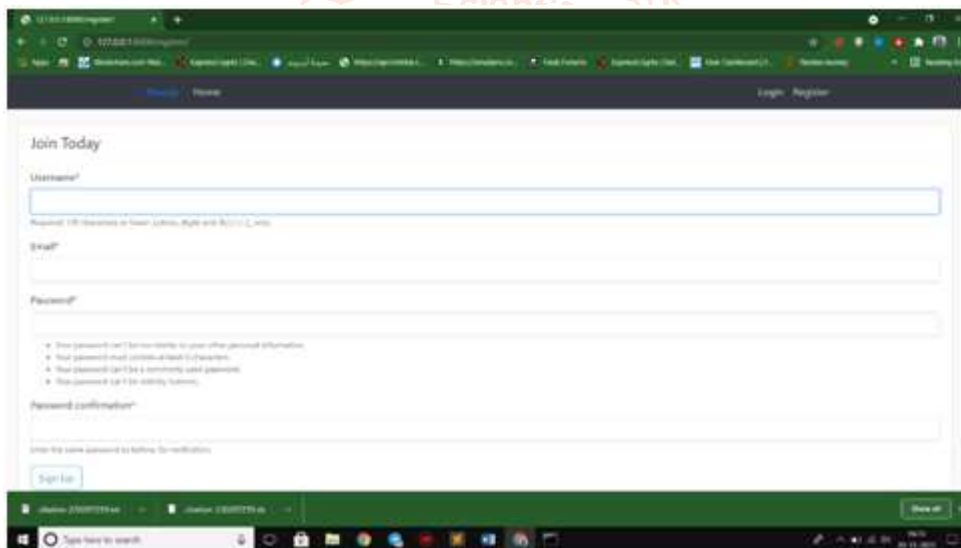


Fig 4: Sign-up

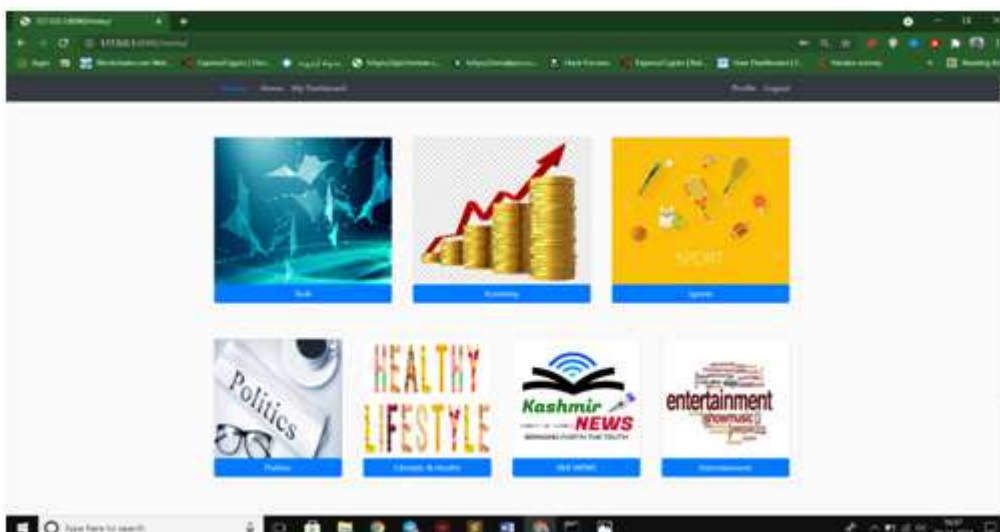


Fig 5: Dashboard

6. CONCLUSION

Using news aggregation is one of the best ways to stay on top of the news and topics you want. They offer convenience and time-saving features. You don't have to invest time in separate content. A content aggregator also reduces all the efforts that are otherwise to be put into data analyzing, demonstrating information, and creating engaging content. Hence, it benefits in saving a lot of time and money.

7. REFERENCES

- [1] www.google.com
- [2] <https://w3schools.com>
- [3] www.stackoverflow.com
- [4] <https://www.simplilearn.com/tutorials/python-tutorial>

